MOLECULAR AND CELLULAR BIOLOGY, Aug. 1992, p. 3365–3371
0270-7306/92/083365-07$02.00/0

Vol. 12, No. 8

Reexamination of Gene Targeting Frequency as a Function of the Extent of Homology between the Targeting Vector and the Target Locus

CHUXIA DENG AND MARIO R. CAPECCHI*

Howard Hughes Medical Institute, Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84112

Received 23 March 1992/Accepted 6 May 1992

Mutations were targeted to the *Hprt* locus of mouse embryo-derived stem cells by using 22 different sequence replacement and sequence insertion vectors. The targeting frequency was examined at two sites within the *Hprt* locus as a function of the extent of homology between the targeting vector and the target locus. The targeting frequency was also compared by using vectors prepared from isogenic and nonisogenic DNA sources. With one exception, all of the vectors showed the same exponential dependence of targeting efficiency on the extent of homology between the targeting vector and the target locus. This was true regardless of whether they were sequence replacement or sequence insertion vectors, whether they were directed toward either of the two different sites within the *Hprt* locus, or whether they were prepared from isogenic or nonisogenic DNA sources. Vectors prepared from isogenic DNA targeted four to five times more efficiently than did the corresponding vectors prepared from nonisogenic DNA. The single case of unexpectedly low targeting efficiency involved one of the vectors prepared from nonisogenic DNA and could be attributed to an unfavorable distribution of heterology between the *Hprt* sequences present in the targeting vector and the endogenous *Hprt* gene.

Gene targeting in mouse embryo-derived stem (ES) cells provides the means for modifying, in a defined manner, virtually any gene in the mouse (2, 3). Analysis of the phenotypes of mice carrying such mutations has already provided insight into the biological role of more than 30 genes whose function had previously been only a matter of speculation.

The first step in undertaking a gene targeting experiment is to design a vector that will deliver the desired mutation to the genome. Some of the concerns which face the investigator are (i) whether to use sequence replacement or sequence insertion vectors, (ii) how much sequence homologous to the target locus to include on the targeting vector, (iii) what advantage is gained by preparing the vector from isogenic DNA, and (iv) whether it is more efficient to use screening procedures (6) or enrichment procedures (8) to identify cell lines containing the desired targeting event. In an earlier publication, we suggested that the gene targeting frequency was strongly dependent on the extent of homology between the targeting vector and the target locus (15). A 2-fold increase in the extent of homology resulted in a 20-fold increase in the gene targeting frequency. Further, sequence replacement and sequence insertion vectors containing equivalent amounts of *Hprt* sequence yielded similar targeting frequencies. However, that study was limited to the analysis of five vectors (three replacement and two insertion) directed at a single site within the *Hprt* locus. Since the time of that publication, it has been suggested that although the targeting frequency is dependent on the extent of homology between targeting vector and target locus, the recombination system saturates at a relatively low extent of homology (i.e., 6 kb [4]) and that sequence insertion vectors recombine much more efficiently than do sequence replacement vectors (5). These reports prompted us to examine the effect of these parameters on the gene targeting frequency in greater depth. Here we report results obtained by using 22 vectors (13 sequence replacement and 9 sequence insertion) directed at two sites, exons 3 and 8, within the *Hprt* locus. These two exons are separated by approximately 18 kb of genomic sequence and provided us an opportunity to determine whether there are local differences in the targeting frequency within the same gene. Further, to assess the degree to which DNA sequence heterology between strains of mice significantly affects the gene targeting frequency, we tested targeting vectors prepared from DNA isolated either from the strain of mouse from which the recipient ES cell line was derived (isogenic DNA) or from a different strain of mouse (nonisogenic DNA).

MATERIALS AND METHODS

Vectors. The targeting vectors used in this study contain the neomycin cassette pMC1Neo (15) inserted into either exon 3 or exon 8 of *Hprt*. The targeting vectors prepared from isogenic DNA were constructed from *Hprt* genomic sequence isolated from a λDASHII-CC1.2 genomic library (prepared by Derrick E. Rancourt). The exon 3 and exon 8 vectors prepared from nonisogenic DNA used *Hprt* inserts derived from λZAPII-Swiss Webster and λCharon 4A-BALB/c genomic libraries, respectively. The structures of the targeting vectors are shown in Fig. 1 and 2. The number associated with each vector indicates the amount of *Hprt* genomic sequence, in kilobases, present in the vector. The suffixes I and nI indicate whether the vector was prepared from isogenic (I) or nonisogenic (nI) DNA. To prepare the isogenic replacement vector illustrated in Fig. 1, an 18.6-kb *Hprt* fragment containing exons 2 and 3, and derived from the λDASHII-CC1.2 genomic library, was subcloned into the plasmid vector T7/T3α19 (Bethesda Research Laboratories). To construct pRV18.6-I, pMC1Neo was inserted into the *Xho*I site of exon 3. The transcriptional direction of *neo* was the same as that of

* Corresponding author.

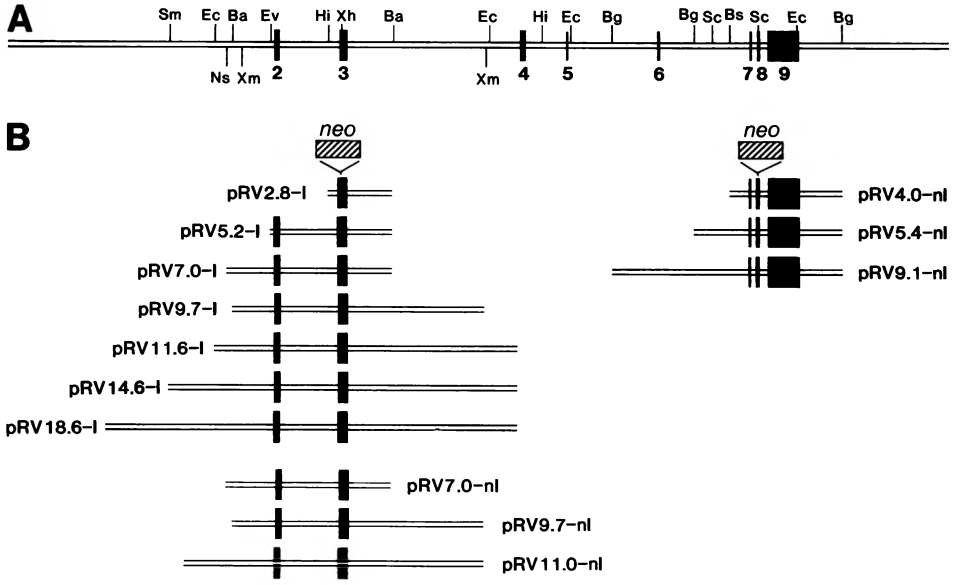

FIG. 1. Maps of the sequence replacement vectors. The number associated with each targeting vector denotes the amount of *Hprt* genomic sequence in kilobases present in the vector. Each of the vectors contains the neomycin cassette, pMC1Neo, inserted into either exon 3 or exon 8. I denotes that the vector was made from isogenic CC1.2 DNA; nI denotes that the vector was prepared from nonisogenic DNA, Swiss Webster DNA for the exon 3 vectors and BALB/c DNA for the exon 8 vectors. Restriction endonuclease sites: Ba, *Bam*HI; Bg, *Bg*lII; Bs, *Bst*EII; Ec, *Eco*RI; Ev, *Eco*RV; Hi, *Hin*dIII; Ns, *Ns*iI; Sc, *Sca*I; Sm, *Sma*I; Xh, *Xho*I; Xm, *Xmn*I.

Hprt. The replacement vectors pRV2.8-I through pRV14.6-I were derived from pRV18.6-I; their maps are illustrated in Fig. 1. Construction of pRV4.0-nI, pRV5.4-nI, and pRV9.1-nI has been previously described (15). pRV11.0-nI was prepared by subcloning an 11.0-kb *Eco*RI-*Xmn*I *Hprt* fragment, containing exons 2 and 3 and derived from the Swiss Webster genomic library, into the plasmid vector T7/T3α19 and again inserting pMC1Neo into the *Xho*I site of exon 3. The direction of transcription of *neo* was again the same as that of *Hprt*. pRV7.0-nI and pRV9.7-nI were derived from pRV11.0-nI. The insertion vectors illustrated in Fig. 2 were constructed in a variety of ways. Construction of pIV3.7-nI and pRV9.3-nI has been previously described (15). pIV7.0-I was constructed by subcloning a 7-kb *Bam*HI fragment (obtained from the 18.6-kb fragment previously described) containing exons 2 and 3 into the T7/T3α19 plasmid vector and then inserting pMC1Neo into the *Sal*I site in the polylinker of T7/T3α19. pIV6.8-nI contains an *Eco*RI *Hprt* fragment and was constructed in such a way that it can be used as either an insertion or replacement vector, depending on the restriction endonuclease sites used to linearize the vector. The *Hprt* sequences used to make pIV3.0-nI, pIV7.4-nI, pIV12.1-nI, and pIV14.3-nI were prepared from the Swiss Webster genomic library. Each specific *Hprt* fragment was circularized, digested with *Bst*EII, and cloned at the *Bst*EII site of pUC9, after which pMC1Neo was inserted into exon 8 of *Hprt*. Linearization of the targeting vector with *Bst*EII releases pUC9 and generates the corresponding insertion vector.

Isolation of Hprt$^-$ targeted cell lines. The experiments used the mouse ES cell line CC1.2 (1). Linearized vector DNA was transfected by electroporation into CC1.2 cells. For a targeting vector that contains 10 kb of DNA sequence, the concentration of DNA during electroporation was 25 μg/ml. The DNA concentration for any specific vector was adjusted so that the number of DNA molecules per cell during electroporation was the same. Just prior to electroporation, the ES cells were trypsinized, washed once with serum-free Dulbecco modified Eagle medium, and resuspended in electroporation buffer (20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES], pH 7.0, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose, 0.2 mM β-mercaptoethanol) at a concentration of 10^7 cells per ml. DNA was added to 1 ml of the cell suspension, mixed, and given two electrical pulses at 625 V/cm (15 μF) with a Promega Biotec X Cell 2000 electroporator. Under these conditions, 50% ± 10% of the cells survived electroporation. Following electroporation, the cells were diluted with Dulbecco modified Eagle medium plus serum and plated on mitomycin-treated Neo^r STO feeder cells at a concentration of 10^6 ES cells per 100-mm-diameter petri dish. The cells were fed with fresh medium every other day. After 24 h of incubation at 37°C in a CO_2 incubator, the medium was supplemented with G418 (250 μg/ml). The cells were maintained under G418 selection for 9 days, followed by 6 days of selection in G418 plus 6-thioguanine (6-TG; 1 μg/ml). At the end of this period, the number of surviving colonies was determined. The targeting frequencies in Fig. 4 and 5 are expressed as the number of G418^r 6-TG^r colonies per number of cells surviving electroporation (i.e., ~50% of the cells that were electroporated). Each point on the curves shown in Fig. 4 and 5 represents the pooling of data from two or more electroporation experiments. Following the introduction of the *Hprt* targeting vectors, the DNA from over 600 independent G418^r 6-TG^r cell lines was analyzed by Southern transfer analysis, and in each case we found that the loss of *Hprt* gene function was due to targeted disruption of the *Hprt* gene.

RESULTS

The *Hprt* gene was chosen as the locus for these experiments because cell lines containing targeted disruptions of this gene can be selected directly. The *Hprt* gene resides on the X chromosome, and therefore the ES cells, which have

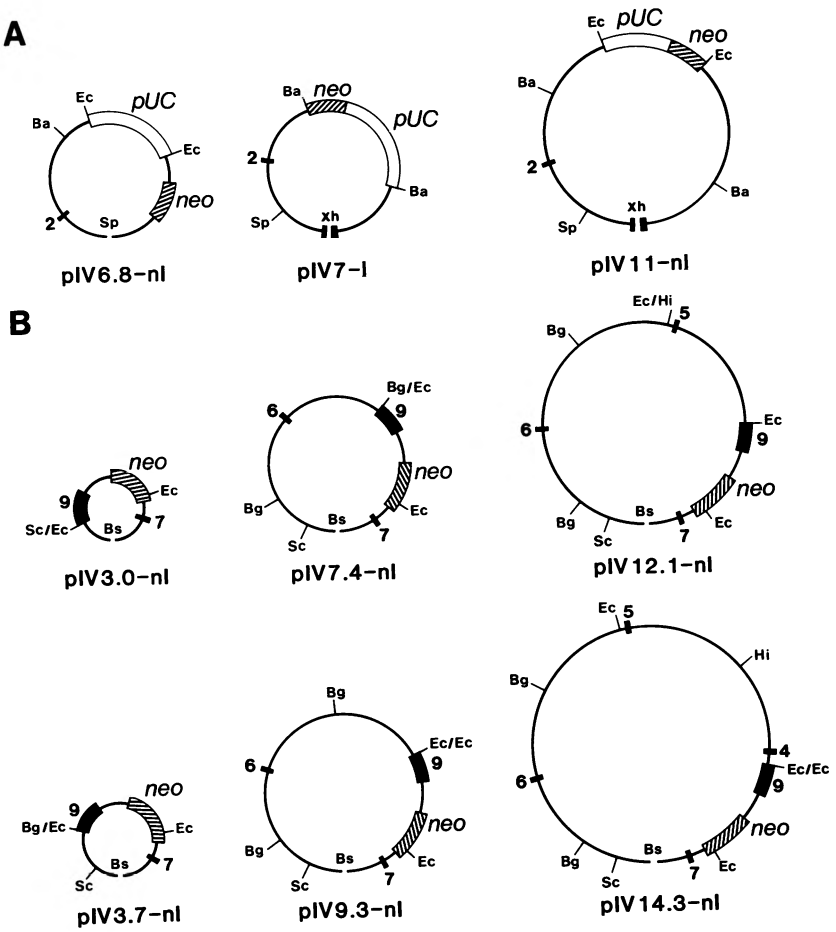

FIG. 2. Maps of the sequence insertion vectors. (A) Sequence insertion vectors directed at disrupting the *Hprt* gene at exon 3; (B) sequence insertion vectors directed at disrupting the *Hprt* gene at exon 8. The targeting vector nomenclature and restriction enzyme sites are described in the legend to Fig. 1. Other restriction endonuclease site: Sp, *Spe*I.

a male karyotype, contain only a single copy of this gene. Inactivation of *Hprt* renders the cells resistant to the guanine analog 6-TG, which kills Hprt$^+$ cells. The *Hprt* targeting vectors contain the *neo* gene inserted into either exon 3 or exon 8 of *Hprt* (Fig. 3). The *neo* gene performs two functions: it disrupts *Hprt* coding sequences in the vector, and it confers resistance to the cytotoxic drug G418 on cells that have acquired a functional copy of the gene, by either homologous or random insertion. Following transfection with the targeting vector, the probability that a cell acquires a random insertion of the targeting vector and concomitantly acquires an independent, spontaneous null mutation in the *Hprt* gene is so low, relative to the *Hprt* target disruption frequency, that such cell lines do not present a background problem (see Materials and Methods). Following selection for G418^r 6-TG^r, the only cell lines that we have obtained are those that became Hprt$^-$ by targeted disruption.

Figure 3 illustrates the use of sequence replacement and sequence insertion vectors. Sequence replacement and sequence insertion vectors can contain identical DNA sequences but are topologically distinct. In sequence replacement vectors (illustrated at exon 3 of *Hprt*), both sides of the mutation (the *neo* gene) are flanked by genomic sequences colinear with the target locus. Following pairing between the vector and genomic sequences, homologous recombination by either gene conversion or double-reciprocal recombination results in the replacement of endogenous sequences by exogenous sequences. An important point for later reference is that the transfer of the mutation by using a replacement vector requires crossovers to occur on both sides of the mutation. Sequence insertion vectors contain a double-strand break within sequences homologous to the target locus (illustrated at exon 8 of *Hprt* in Fig. 3). Thus, the termini of the linearized vector lie adjacent to one another on the *Hprt* map. Pairing of these vectors with their genomic homolog followed by recombination at or near the double-strand break results in an insertion of the entire vector into the endogenous gene. This produces a null mutation, since a contiguous functional gene is no longer present in these cells. However, insertion of the vector into the target locus results in a partial duplication of genomic *Hprt* sequences, and therefore the target locus is susceptible to reversion to Hprt$^+$ through a secondary intramolecular homologous recombination event mediated by the duplicated sequences.

In this study, the influences of four variables on the gene targeting frequency were examined: (i) the type of vector, sequence replacement, and sequence insertion used, (ii) the amount of genomic *Hprt* sequence included in each vector, (iii) the site within a gene chosen for targeting, here either exon 3 or exon 8 of *Hprt*, and (iv) the degree of homology between the target and the source of vector DNA, with the source here being mice either isogenic or nonisogenic to the

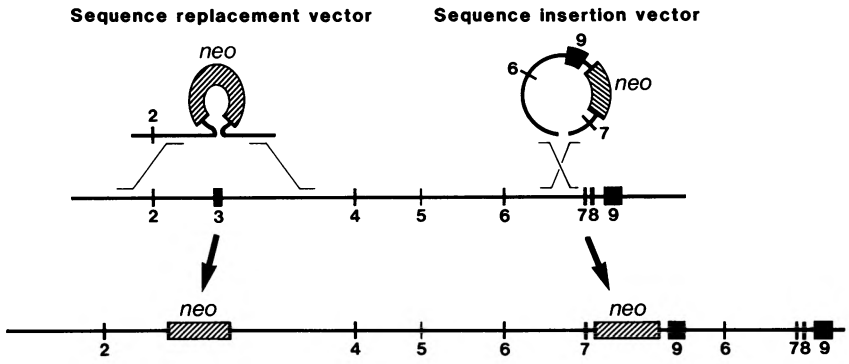

FIG. 3. Disruption of the *Hprt* gene by sequence replacement and sequence insertion vectors. Two schemes for gene disruption of the *Hprt* gene are depicted, one at exon 3 by sequence replacement and a second at exon 8 by sequence insertion. In sequence replacement vectors, the mutation, in this case the *neo* gene inserted into exon 3, is flanked on both sides by *Hprt* genomic sequence which is colinear with the endogenous *Hprt* gene. A recombination event on both sides of the mutation, gene conversion or double-reciprocal exchange, results in the replacement of endogenous sequence with the exogenous, *neo*-containing sequence. With the sequence insertion vector, illustrated at exon 8, a double-strand break in *Hprt* sequences places the ends of the linearized vector adjacent to one another on the *Hprt* map. A recombination event at or near the double-strand break results in insertion of the entire vector into the endogenous *Hprt* locus. This produces a duplication of a portion of the *Hprt* gene at the endogenous locus.

targeted cell strain. Figures 1 and 2 provide a summary of the structures of the targeting vectors used in this study. The sequence replacement (Fig. 1) and sequence insertion (Fig. 2) vectors are directed toward disrupting either exon 3 or exon 8. Disruptions at either exon resulted in tight null mutations. Following targeted disruption of the endogenous *Hprt* locus in ES cells, *Hprt* activity could not be detected either by enzyme assays in cell extracts or by incorporation of [^{3}H]hypoxanthine in intact cells.

Figure 4 illustrates the absolute targeting frequency plotted as a function of the length of nonisogenic *Hprt* genomic sequences present in the targeting vector. Included on these plots are results obtained with sequence replacement or sequence insertion vectors directed toward disrupting either exon 3 or exon 8. Except for the results of one vector, the targeting frequency shows a very strong dependence on the length of homology between the targeting vector and the target locus, irrespective of whether the vector is of the sequence replacement or sequence insertion type. The targeting frequency is the same at exon 3 as it is at exon 8. In the range between 2 and 10 kb, this dependency is exponential, yielding a 100-fold increase in the gene targeting frequency. The recombination system appears to saturate with respect to length of homology at ~14 kb (Fig. 4).

There is one interesting exception to these observations: the sequence replacement vector, pRV7.0-nI, behaved anomalously. This targeting vector was 10-fold less efficient than predicted from the curve (Fig. 4). However, the corresponding sequence insertion vector behaved as expected (Fig. 4). It is worth noting that the marked difference in targeting efficiency of this particular replacement vector, relative to the corresponding insertion vector, can be eliminated by extending the amount of *Hprt* sequence in both vectors by 4 kb. Thus, pRV11.0-nI and pIV11.0-nI showed very similar gene targeting frequencies (Fig. 4).

There are a number of compelling reasons that targeting vectors should be prepared from DNA isolated from the same mouse strain as that from which the recipient ES cell line was derived (isogenic DNA), and indeed for a number of years we have been preparing many of our targeting vectors from CC1.2 DNA. In bacteria, exchange of information between homologous regions of DNA is edited by mismatch

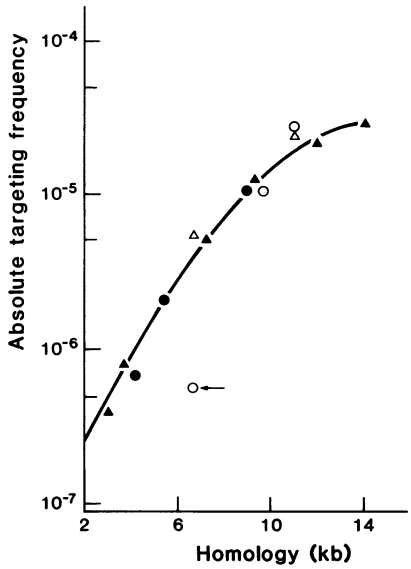

FIG. 4. The targeting frequency at the *Hprt* locus as a function of the extent of homology between the targeting vector and the endogenous target. All of the vectors used for this study were prepared from nonisogenic DNA. The vectors were directed at disrupting either exon 3 (open symbols) or exon 8 (filled symbols). Included are data obtained by using either sequence replacement vectors (circles) or sequence insertion vectors (triangles). Each point represents the pooling of data from two or more independent electroporation experiments (see Materials and Methods). Plotted are the number of targeted Hprt$^-$ cell lines (i.e., G418^r 6-TG^r cell lines) obtained following transfection with a given targeting vector per number of cells electroporated × 0.5 (i.e., the number of cells surviving electroporation). One vector, indicated by the arrow and designated pRV7.0-nI, yielded a gene targeting frequency much lower than that predicted from the curve. We attribute this anomalous behavior to an unfavorable distribution of heterology between the targeting vector and the endogenous *Hprt* gene (see text). The number of targeting events per total number of ES cells surviving electroporation is plotted on a logarithmic scale as a function of the number of kilobases of *Hprt* genomic sequence contained within each of the targeting vectors.

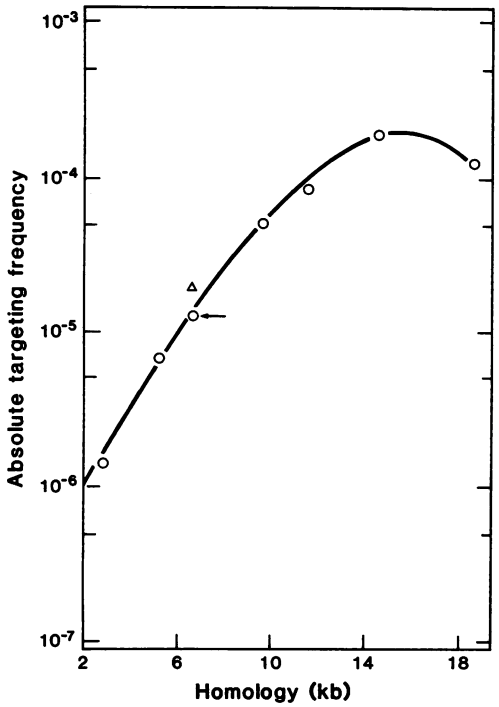

FIG. 5. Targeting frequency with use of isogenic DNA (CC1.2) as a function of the extent of homology. All of the vectors are directed at disrupting exon 3 of *Hprt*. Circles represent sequence replacement vectors; the triangle represents an insertion vector. Note that the replacement vector pRV7.0 (arrow), when prepared from isogenic DNA, no longer behaves aberrantly.

repair proteins whose function is to abort or inactivate heteroduplex intermediates containing excessive numbers of mismatched base pairs (10, 11). It would be surprising if higher eukaryotes, whose genomes are laden with closely related repetitive DNA elements, did not contain similar editing machinery. Further, Letsou and Liskay have shown that the frequency of intrachromosomal gene conversion can be very sensitive to even a single base pair mismatch within a 1-kb interval (7). Most recently, teRiele et al. have shown that the targeting frequency at the retinoblastoma locus is 10- to 20-fold higher when the targeting vector is prepared from isogenic rather than nonisogenic DNA (14).

The data in Fig. 4 show a strong dependency of the gene targeting frequency on the extent of homology between targeting vector and target locus for vectors prepared from nonisogenic DNA. This dependency might thus reflect the need for a minimal stretch of DNA in the targeting vector with perfect homology to the target locus; the longer the vector, the greater the probability for the presence of such a perfect match. If this were the case, a very different relationship between the gene targeting frequency and the length of homology might be expected for vectors prepared from isogenic DNA, in which the match would be perfect along the entire length of the *Hprt* genomic sequence. Figure 5 shows a plot of the relationship obtained with vectors prepared from isogenic DNA. The shape of the curve is remarkably similar to that shown in Fig. 4. Between 2 and 10 kb, the relationship between the length of homology and the frequency of targeting events is again exponential. The recombination system again appears to saturate with respect to this parameter at ~14 kb of homology. Instead of a change

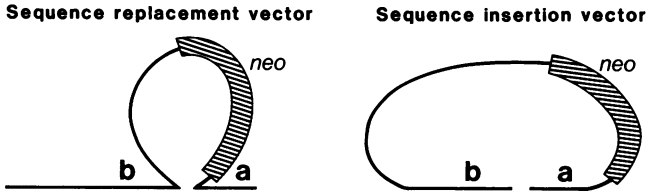

FIG. 6. Schematic comparison of sequence replacement vectors with sequence insertion vectors.

in the shape of the curve, we observe a displacement of the curve such that, on average, vectors prepared from isogenic DNA are four to five times more efficient than the corresponding vectors prepared from nonisogenic DNA. The replacement vector, pRV7.0, when prepared from isogenic DNA, no longer behaved anomalously (arrow in Fig. 5).

DISCUSSION

The effects on the gene targeting frequency were assessed for a number of experimental parameters. For the most part, sequence replacement and sequence insertion vectors behaved equivalently with respect to the targeting efficiency. With both types of vector, we found a strong dependency of the targeting frequency on the length of homology between the targeting vector and the target locus. There was one interesting exception: pRV7.0-nI was 10-fold less efficient than its corresponding insertion vector. We attribute the low targeting efficiency of pRV7.0-nI to an unfavorable distribution of heterology between the vector, derived from nonisogenic sources, and the target locus. Consistent with this interpretation is the finding that when this vector is prepared from isogenic DNA, it no longer behaves aberrantly. Further, as an indicator of extensive heterology in this region of the *Hprt* gene, it was found that 5 of the 30 *Dde*I restriction enzyme cleavage sites within this 7-kb interval were polymorphic when *Hprt* sequences in pRV7.0-nI and pRV7.0-I were compared. From this observation, a conservative estimate is that the base pair heterology between these two sequences exceeds 1%. Though pRV7.0-nI behaved anomalously, the comparable sequence insertion vector did not. That is, the sequence insertion vector appears to be less sensitive to the presence of the same base pair heterology. A number of factors might account for this difference. For example, let us assume that the unfavorable distribution of heterology was on arm a of the vector (Fig. 6). For a sequence replacement vector to transfer the mutation (i.e., the *neo* gene) to the genome, crossovers must occur in both arms a and b. If the heterology reduces the efficiency of crossovers in arm a, then the efficiency of recombination of the entire vector is affected. However, with an insertion vector, though initiation of recombination may occur in arms a and b, a crossover in arm a is not required. A single crossover in any region of homology is sufficient to insert the vector into the endogenous locus. In addition or alternatively, arm b in an insertion vector, because it has *Hprt* sequence contiguous with arm a, may be more effective in stabilizing the pairing of the heterologous arm a with the endogenous locus and thereby enhance the efficiency of recombination. Consistent with this line of reasoning is the observation that the two vector types show the same targeting efficiency when 4 kb of additional *Hprt* sequence is added to each of them.

The same factors that make sequence replacement vectors

potentially more sensitive to unfavorable distribution of heterology would also make replacement vectors more sensitive to having one short arm. If the mutation in a sequence replacement vector is flanked by 1 kb or less of genomic target sequence, then we observe a reduction in the expected targeting frequency as well as a propensity for unpredictable rearrangements at the target locus (16).

Except for events mediated by pRV7.0-nI, targeted recombination at the 3′ and 5′ ends of the *Hprt* gene appears to occur with equal frequency. Within this locus, aside from localized differences in the distribution of heterology, there do not appear to be sequences that markedly enhance or depress targeted recombination. The efficiency of recombination appears to be dependent only on the length of *Hprt* sequence present in the targeting vector. Perhaps unexpectedly, targeting vectors prepared from isogenic DNA show a similar dependence of targeting frequency on length of homology between the targeting vector and the endogenous target. In the *Hprt* locus, isogenic vectors are four- to fivefold more efficient than those of equal length prepared from nonisogenic DNA.

The important parameter for determining the targeting frequency is the length of homology between the targeting vector and the target sequence and not the total length of DNA sequence present in the targeting vector. For example pRV5.2 and pRV7.0 target at the same frequency whether or not the bacterial plasmid sequence is removed from the targeting vector prior to its introduction into ES cells. The presence of the additional 3 kb of nonhomologous DNA at the end or ends of the replacement vector appears to be neutral with respect to its targeting efficiency. Further, similar *Hprt* replacement vectors transferring 8 bp or 1, 3.4, 4.3, or 12 kb of heterologous DNA to the endogenous *Hprt* locus target at similar frequencies (9).

The exponential dependence of targeting frequency on the length of homology between vector and target sequences observed in mammalian cells is much stronger than that observed in bacteria or yeast, for which a linear dependency has been reported (12, 13). It is also curious that the mammalian homologous recombination machinery appears to require ~14 kb of homologous sequence in the targeting vector to saturate its need for extended homology. This feature was common to sequence replacement and sequence insertion vectors as well as to vectors prepared from isogenic and nonisogenic DNA. The need for so much DNA to saturate the recombination machinery also appears to be unique to the homologous recombination system present in higher eukaryotes. Both the exponential dependence on the length of homology and the need for long homologous stretches to saturate the system strongly disfavor exchanges between short, relative to long, homologous DNA sequences. Since repetitive DNA elements in higher eukaryotes are usually less than 1 kb in length, these features of the homologous recombination machinery may have evolved to suppress recombination among repetitive DNA elements, thereby lending stability to the genome.

That 14 kb of homologous DNA sequence is required to saturate the mammalian recombination machinery may be an underestimate. As the total length of the targeting vector exceeds 14 kb, the efficiency of introducing the vector by electroporation into ES cells, as well as its efficiency to diffuse into the nucleus, may decrease. However, we have not observed a decrease in the efficiency of transfection (i.e., no. of G418^r colonies per no. of ES cells electroporated), using vectors whose lengths increased from 10 to 25 kb.

In light of the discussion presented above, the results of Hasty et al. (5) are reconcilable with the data presented herein. They reported that their *Hprt* replacement vector, RV6.8, targeted to the *Hprt* locus at a ninefold-lower frequency than did their comparable insertion vector and that most of the targeting events mediated by this replacement vector yielded unpredictable rearrangements at the target locus (5). Their targeting vectors were prepared from the same region of *Hprt* whose sequence we found to be exceptionally heterologous among different strains of mice. In our own hands, the replacement vector pRV7.0-nI prepared from the same region of *Hprt* and prepared from nonisogenic DNA was 10-fold less efficient than its comparable insertion vector. In addition, the design of their replacement vectors often differed from the design of our replacement vectors in one important parameter. In their replacement vectors, the length of one of the arms of *Hprt* sequence flanking the *neo* gene was 1 kb or less, whereas in our replacement vectors, the shortest arm was usually several kilobases. In a separate report (16), we have shown that replacement vectors in which the *neo* gene is flanked on both sides with several kilobases of DNA homologous to the target locus mediate replacement of chromosomal sequences with precision, whereas if the length of one of the arms of the vector is reduced to below 1 kb, the fidelity of homologous recombination at the target locus is greatly reduced.

In summary, we have shown that sequence replacement and sequence insertion vectors behave similarly with respect to their targeting efficiencies in the mouse. When a targeting vector behaves anomalously, an unfavorable distribution of heterology between the targeting vector and the target locus may be the culprit.

ACKNOWLEDGMENTS

We thank Marjorie Allen, Sheila Campero, and Carol Lenz for excellent technical assistance.

REFERENCES

1. **Bradley, A., M. Evans, M. H. Kaufman, and E. Robertson.** 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature (London) **309:**255–256.
2. **Capecchi, M. R.** 1989. The new mouse genetics: altering the genome by gene targeting. Trends Genet. **5:**70–76.
3. **Capecchi, M. R.** 1989. Altering the genome by homologous recombination. Science **244:**1288–1292.
4. **Hasty, P., J. Rivera-Perez, and A. Bradley.** 1991. The length of homology required for gene targeting in embryonic stem cells. Mol. Cell. Biol. **11:**5586–5591.
5. **Hasty, P., J. Rivera-Perez, C. Chang, and A. Bradley.** 1991. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. Mol. Cell. Biol. **11:**4509–4517.
6. **Kim, H. S., and O. Smithies.** 1988. Recombinant fragment assay for gene targetting based on the polymerase chain reaction. Nucleic Acids Res. **16:**8887–8903.
7. **Letsou, A., and R. M. Liskay.** 1987. Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells. Genetics **117:**759–769.
8. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the protooncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to nonselectable genes. Nature (London) **336:**348–352.
9. **Mansour, S. L., K. R. Thomas, C. Deng, and M. R. Capecchi.** Introduction of a *lacZ* reporter gene into the mouse *int-2* locus by homologous recombination. Proc. Natl. Acad. Sci. USA **87:**7688–7692.
10. **Rayssiguier, C., D. S. Thaler, and M. Radman.** 1989. The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. Nature (London) **342:**396–401.

11. **Shen, P., and H. V. Huang.** 1989. Effect of base pair mismatches on recombination via the RecBCD pathway. Mol. Gen. Genet. **218:**358–360.
12. **Smolik-Utlant, S., and T. D. Petes.** 1983. Recombination of plasmids into the *Saccharomyces cerevisiae* chromosome is reduced by small amounts of sequence heterogeneity. Mol. Cell. Biol. **3:**1204–1211.
13. **Szostak, J. W., and R. Wu.** 1979. Insertion of a genetic marker into the ribosomal DNA of yeast. Plasmid **2:**536–554.
14. **teRiele, H., E. R. Maandag, and A. Berns.** Highly efficient targeting in embryonic stem cell via homologous recombination with isogenic DNA constructs. Proc. Natl. Acad. Sci. USA, in press.
15. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.
16. **Thomas, K. R., C. Deng, and M. R. Capecchi.** 1992. High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. Mol. Cell. Biol. **12:**2919–2923.